BMC Genomics

RESEARCH

Open Access

Genome-wide characterization of regulator of chromosome condensation 1 (*RCC1*) gene family in *Artemisia annua* L. revealed a conservation evolutionary pattern

Jieting Chen[1†], Wenguang Wu[2†], Xiaoxia Ding[1], Danchun Zhang[1], Chunyan Dai[1], Hengyu Pan[1], Peiqi Shi[1], Chanjuan Wu[3], Jun Zhang[3], Jianmin Zhao[3], Baosheng Liao[1*], Xiaohui Qiu[1*] and Zhihai Huang[1*]

Abstract

Background *Artemisia annua* is the major source for artemisinin production. The artemisinin content in *A. annua* is affected by different types of light especially the UV light. *UVR8*, a member of *RCC1* gene family was found to be the UV-B receptor in plants. The gene structures, evolutionary history and expression profile of *UVR8* or *RCC1* genes remain undiscovered in *A. annua*.

Results Twenty-two *RCC1* genes (*AaRCC1*) were identified in each haplotype genome of two diploid strains of *A. annua,* LQ-9 and HAN1. Varied gene structures and sequences among paralogs were observed. The divergence of most *RCC1* genes occurred at 46.7 – 51 MYA which overlapped with species divergence of core Asteraceae during the Eocene, while no recent novel *RCC1* members were found in *A. annua* genome. The number of *RCC1* genes remained stable among eudicots and *RCC1* genes underwent purifying selection. The expression profile of *AaRCC1* is analogous to that of *Arabidopsis thaliana (AtRCC1)* when responding to environmental stress.

Conclusions This study provided a comprehensive characterization of the *AaRCC1* gene family and suggested that *RCC1* genes were conserved in gene number, structures, constitution of amino acids and expression profiles among eudicots.

Keywords *Artemisia annua*, *RCC1*, UVR8, UV-B, Purifying selection

† Jieting Chen and Wenguang Wu contributed equally to this work.

*Correspondence:
Baosheng Liao
liaobaosheng@gzucm.edu.cn
Xiaohui Qiu
qiuxiaohui@gzucm.edu.cn
Zhihai Huang
zhhuang7308@163.com
[1] Key Laboratory of Quality Evaluation of Chinese Medicine of the Guangdong Provincial Medical Products Administration, the Second Clinical College, Guangzhou University of Chinese Medicine, Guangzhou 510006, China
[2] Artemisinin Research Center, China Academy of Chinese Medical Sciences, Beijing 100700, China
[3] Sunribio Co.Ltd, Shenzhen 518101, China

Background

Artemisia annua, a traditional Chinese medicine, belonging to Asteraceae family, is the major source for artemisinin which is widely used in the treatment of malaria [1, 2]. Artemisinin-based combination therapies (ACTs) have been highly recommended by the World Health Organization for treating malaria [3–5]. Though semi-synthetic artemisinin has been developed [6, 7], low yield and high cost make large-scale industrial applications unavailable [8]. Currently, *A. annua* is the major source of artemisinin.

The main distribution areas of *A. annua* were concentrated in mid-latitudes in southeastern Asia, western and central Europe, south-eastern North America and south-eastern South America [9]. In China, *A. annua* grown in the south of the Qinling Mountains-Huaihe River Line had a higher artemisinin content compared to the northern ones [10]. Humidity and sunshine duration were speculated as major limiting ecological factors that affect the accumulation of artemisinin [11]. *A. annua* is a determinate short-day plant with a critical photoperiod [12], while biomass and artemisinin production were increased in response to long-day photoperiod [13]. The daylight contains a variety of radiation, of which the Ultraviolet-B radiation (UV-B, 280–315 nm) [14] is an important environmental signal that pleiotropically regulates development, morphogenesis and physiology in plants [15]. Previous studies have demonstrated that short-term UV-B treatment to *A. annua* may be a safe approach to accumulating artemisinin content while acting on stress-regulated genes to keep the plant healthy [16]. Besides, UV-B radiation and phytohormone gibberellins coordinately promoted the accumulation of artemisinin in *A. annua*, with a significant up-regulation of two genes in artemisinin biosynthetic pathway (*ADS* and *CYP71AV1*) [17].

UV RESISTANCE LOCUS 8 (UVR8) is an evolutionarily well conserved UV-B photoreceptor that regulates UV-B photomorphogenesis in plants [18], which employs a unique photosensory mechanism for light absorption and initiation of the signaling events that eventually lead to particular physiological responses [19–21]. UVR8 contains sequence similarity and predicted structural similarity to human Regulator of Chromatin Condensation 1 (RCC1), whose sequence is highly conserved among all eukaryotes and consists of a seven-bladed-β-propeller, also known as seven RCC1 repeat units [22, 23]. *RCC1* functions as a guanine-nucleotide-exchange factor (GEF) for the Ran G-protein to regulate diverse biological processes, nucleocytoplasmic transport, and the cell cycle [24]. *UVR8* is a member of *RCC1* gene family, which strongly associates with chromatin, while UVR8 has little Ran GEF activity and it is present in both the cytosol and nucleus in contrast to other *RCC1* family proteins localized in the nucleus [25]. Normally, UVR8 is evenly distributed in the cytoplasm and nucleus, however, under UV-B treatment, it tends to accumulate in the nucleus through interaction with constitutive photomorphogenic 1 protein (COP1), triggering a UV-B cascade [18, 26, 27]. The amino acid sequence of UVR8 is enriched with aromatic residues [28]. The aromatic amino acids refer to amino acids with benzene ring in molecular structure, including tyrosine (F), phenylalanine (P) and tryptophan (W), which is bound up with UV absorption [29]. The *Arabidopsis thaliana* (AtUVR8) has 14 W residues, among which W285 and W233 were shown to have an important role in UV-B-triggered signaling [30, 31]. The participation of UVR8 in the UV-B response is UV-B dose-dependent, which mediates several responses to low doses of UV-B, while high UV-B doses trigger other adaptive mechanisms [32]. The UVR8 and other members in the *RCC1* gene family have been identified in a range of plant species [33, 34]. For instance, *Spartina alterniflora* RCC1 (SaRCC1), negatively regulates salt stress responses by affecting stress-related gene expression [33]. RUG3 (a mitochondrial protein) is required for efficient splicing of the nad2 mRNA, which encodes a complex I subunit in mitochondria of *A. thaliana* [35]. Tolerant to Chilling and Freezing 1 protein (TCF1), interacts with histones H3 and H4 and associates with chromatin containing a target gene, encoding a glycosylphosphatidylinositol-anchored protein that regulates lignin biosynthesis, and thus affect the freezing tolerance of plants [36, 37]. RCC1-like domain (RLD) proteins, identified as LZY interactors, are essential regulators of polar auxin transport [38]. SAB1 is a crucial new component of ABA signaling which negatively regulates ABI5 through multidimensional mechanisms during post-germination in *A. thaliana* [39]. Currently, the *UVR8* and *RCC1* gene family in *A. annua* genome has not been reported as well as their evolutionary history.

In this study, a comprehensive bioinformatic analysis was conducted on the *RCC1* gene family at the genome-wide level of four haplotype genomes of two *A. annua* strains, including gene structures, phylogenetic relationship construction, gene variation and gene expression profile, which could provide useful information for further functional investigations of *A. annua*.

Results

Identification and characterization of *AaRCC1* genes

Genes with *RCC1* domain (PF00415) were defined as candidate *RCC1* genes and then manually corrected. In total, 22 *RCC1* genes (named *AaRCC1_01* to *AaRCC1_22*) were identified in each haplotype of *A. annua* (LQ-9 haplotype 0 and haplotype 1, HAN1 haplotype 0 and

haplotype 1) (Fig. 1A and Table S1). *RCC1* gene number was consistent among four haplotypes. Gene structures varied among gene members (Fig. 1B). The exon numbers varied from 4 to 16, and gene length ranged from 2, 911 (*AaRCC1_15*) to 11, 666 (*AaRCC1_19*) bp (Table 1). The RCC1 domain number varied from 4 to 7 and some members had PH (PF00169), BRX (PF08381) or FYVE (PF01363) domain (Fig. 1C). Genes with the same exon number and function domain annotated in *RCC1_14*, *16*, *17*, *18*, *19* clustered in a same clade, showing similarities

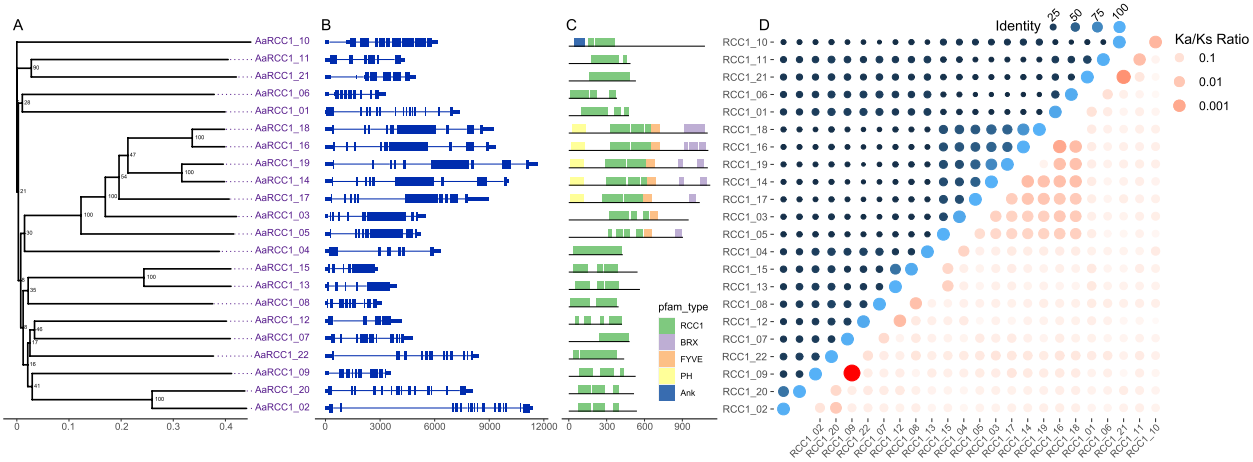

Fig. 1 Characteristics of *AaRCC1* genes. **A** Phylogenetic relationships (numbers on the nodes represent supporting values). **B** Gene structures. Blue rectangles represent the coding sequences, thin blue lines connecting two exons represent introns, and thick blue lines represent 5′-UTR or 3′-UTR. **C** Domain information identified by PfamScan. **D** Identity and Ka/Ks values between alleles

Table 1 The basic information about the *RCC1* genes in *A. annua* LQ-9 haplotype 0

Gene name	Gene length (bp)	CDS length (bp)	Intron/ Exon	Pep length	MW	PI	Tryptophan content (%)	Aromatic amino acid content (%)
AaRCC1_01	7375	1434	14/15	477	51,568.61	6.5	1.89	7.06
AaRCC1_02	11,400	1624	15/16	540	58,375.81	9.37	1.48	9.01
AaRCC1_03	5477	2853	7/8	950	104,348.87	6.92	1.26	6.11
AaRCC1_04	6354	1287	5/6	428	45,206.27	6.64	2.57	7.68
AaRCC1_05	5265	2718	7/8	905	99,733.53	9.19	1.55	7.24
AaRCC1_06	3312	1149	8/9	382	40,711.91	5.6	2.36	7.18
AaRCC1_07	4776	1449	10/11	482	50,596.17	5.62	2.90	6.54
AaRCC1_08	3089	1188	9/10	395	42,267.78	5.85	3.80	7.26
AaRCC1_09	3568	1593	8/8	466	50,337.21	6.15	2.58	6.58
AaRCC1_10	6180	3240	10/10	1079	117,690.82	9.36	1.39	8.80
AaRCC1_11	4340	1470	5/6	489	53,398.99	7.53	1.02	5.75
AaRCC1_12	4192	1269	4/4	422	44,822.19	5.37	2.84	9.41
AaRCC1_13	3921	1695	4/5	536	58,024.84	5.43	2.80	6.16
AaRCC1_14	10,082	3366	8/9	1121	121,840.35	8.65	1.34	8.58
AaRCC1_15	2911	1638	4/5	545	58,757.24	5.12	2.94	6.42
AaRCC1_16	9373	3321	8/9	1106	119,713.11	9.08	1.45	8.99
AaRCC1_17	8934	3117	9/10	1038	114,125.44	8.83	1.25	6.33
AaRCC1_18	9223	3306	8/9	1101	119,920.67	9.11	1.45	7.03
AaRCC1_19	11,666	3309	8/9	1102	120,245.48	8.89	1.45	6.18
AaRCC1_20	8104	1554	14/15	517	54,888.78	7.57	1.55	6.90
AaRCC1_21	4945	1559	6/5	532	57,086.69	5.91	1.13	6.00
AaRCC1_22 (*AaUVR8*)	8141	1320	11/12	439	47,637.46	5.44	2.96	7.33

on gene structures and domain regions. The average content of tryptophan (1.96%) and aromatic amino acids (7.54%) in AaRCC1 proteins were significantly higher than those of other proteins in the whole genome (average tryptophan 1.38%, average aromatic amino acids 3.19%, p-value < 0.05). The AaRCC1_22 showing the highest protein identity (76.52%) to that of *Arabidopsis thaliana* (AtUVR8) was identified as AaUVR8, which has relatively high W content in protein sequences among all *RCC1* genes (Table 1). All CDS and protein sequences of *RCC1* genes were aligned pairwisely. High similarities were detected among alleles of each *AaRCC1* gene (CDS sequence identity 79.61–100%, protein sequence identity 94.20%-100%), while sequence variation existed (protein sequence identity 10.43%-82.11%) (Fig. 1D). Notably, most of the Ka/Ks values calculated between alleles and gene members were less than 1, indicating these *RCC1* genes were under purifying selection (Fig. 1D) and tend to eliminate deleterious mutations and maintain functional stability [40].

RCC1 genes are conserved during speciation in eudicot

A comparison analysis of *RCC1* genes between *A. annua* and four other species was conducted. We found a similar gene number of *RCC1* family in five species. There are 24 *RCC1* members in *A. thaliana*, 22 in *A. annua*, 27 in *H. annuus* (Fig. 2A), 23 in *Chrysanthemum nankingense* and 21 in *Vitis vinifera*. In contrast with gene families like terpene synthase (*TPS*) [41] and UDP-glucuronosyltransferase (*UGT*) [42, 43], the gene number of *RCC1* remained conserved without significant expansion by segmental/tandem duplication or whole genome polyploidization. However, duplication debris of *RCC1* was identified for *AaRCC1_07* and *AaUVR8* in *A. annua* genome and duplicated genes were functionally silenced by corrupting of gene structures (Figure S2). Furthermore, similar codon usage was found among *RCC1* orthologs (Fig. 2B). Protein sequences were conserved among *RCC1* orthologs across the five species. For example, the UVR8 showed high conservation on protein sequences of AtUVR8 (*A. thaliana*), CnUVR8 (*C. nankingense*), HaUVR8 (*H. annuus*), AaUVR8 (*A. annua*) and VvUVR8 (*V. vinifera*), especially on W233, W285 and W337 that related to UV-B response functions [20, 44] (Fig. 2C). Ks values among orthologs in each species were calculated pairwisely, which were enriched at a peak near Ks = 1.46 (Fig. 2D). The *RCC1* genes were diverged at 46.7 to 51 MYA based on Ks and r value from *A. thaliana* [45], which overlapped with the time of most subfamilies of core Asteraceae diverged during Eocene [46]. Few novel *RCC1* genes were identified after Eocene.

AaRCC1 exhibited tissue and treatment specific expression profile

The expression profile of *AaRCC1* genes was examined in different tissues (root, stem, leaf, flower) and different treatments (lights with different wavelengths, including

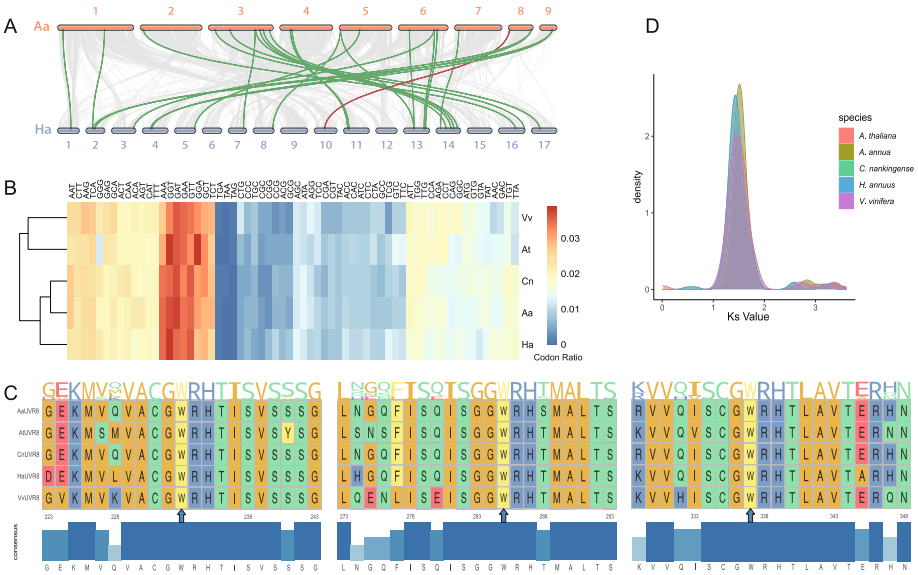

Fig. 2 The feature of RCC1 family in five species. **A** The syntenic relationship of *A. annua* and *H. annuus*, green lines represent syntenic *RCC1* gene pairs (LQ-9 haplotype 0), the red line indicates the *UVR8* gene pair, while the grey background represents other syntenic gene pairs. **B** The codon usage ratio of *RCC1* orthologs in five species. **C** The multiple protein sequence alignment of UVR8 in 5 species. **D** The distribution of *Ks* value of *RCC1* family members in five species

UV-B, blue, red, far-red and white light, phytohormones including gibberellin and brassinolide). The results demonstrated that 22 *RCC1* genes exhibited distinct expression patterns among various conditions (Fig. 3). The expression level of *AaUVR8* in flower was significantly higher than that of other three tissues in two *A. annua* strains (2.4 to 2.7 fold change compared to other tissues, p-value < 0.05). The expression level of *AaRCC1_05* was significantly down-regulated in roots (4.4 to 1.3 fold change compared to other tissues, p-value < 0.05). *AaRCC1_15* and *AaRCC1_17* had higher expression levels in leaves of LQ-9 than those of HAN1. Instead, *AaRCC1_08, AaRCC1_16, AaRCC1_04, AaUVR8 and AaRCC1_07* had higher expression levels in leaves of HAN1, which indicated *RCC1* genes of different strains also showed different expression levels. Expression quantification by qRT-PCR of *AaUVR8* showed 1 to 2.6 fold changes among 18 different strains (Figure S3). Interestingly, after UV-B treatment, five *RCC1* genes including *AaUVR8* showed a decreased gene expression, while *AaUVR8* up-regulated significantly with red light treatment. In contrast, two negative response genes, REPRESSOR OF UV-B PHOTOMORPHOGENESIS 1 (RUP1) and RUP2 [47] were up-regulated after UV-B treatment. A similar expression pattern was detected in *A. thaliana* (Figure S4).

Discussion

RCC1 genes were found to be regulating factors for a series of downstream genes during biological processes, including stress responses under abiotic stress and various hormone treatments [33–39]. Twenty-two *RCC1* genes with significant sequence variations existed in *A. annua* genome and showed different expression patterns in different conditions, which indicated their divergent roles in response to the external environment. Though divergence existed among individual *RCC1* genes in one species, the gene number of *RCC1* genes remained conserved among eudicots. The conservation of the number of *RCC1* genes was not only observed in eudicots but also monocots [48]. Increased gene numbers of *RCC1* could be observed in genomes with recent whole genome polyploidizations [34, 49]. While, the most recent WGT of *A. annua* occurred at 58.12 Ma [50], and most of duplicated *RCC1* copies were functionally inactive with incomplete structures (Figure S2). During the re-diploidization process of post-WGD, gene deletion occurred due to dosage constraints [51]. Functional copies of *RCC1* genes were maintained over time and the *RCC1* gene loss can be an adaptive evolutionary force facing environmental challenges [45]. A negative selection was observed among *RCC1* genes as the Ka/Ks ratios were prevalently lower than 1 within or between species, which would eliminate deleterious mutations and maintain functional stability of *RCC1* genes [40]. As their conservation characteristics among different lineages of eudicots, studies of functional examination and regulatory mechanism deconstruction should be conducted for each *RCC1* member which could be beneficial for the whole plant research community.

UVR8 is one of the *RCC1* family members and the well-known UV-B receptor gene [22]. UV-B radiation is an environmental stimulus, a major abiotic stress confronting living tissue. Low-dose and non-damaging UV-B regulate photomorphogenesis and metabolite biosynthesis by serving as a photomorphogenic signal [52]. The photoactivated UVR8 could transduce UV-B signal via multiple mechanisms to regulate transcription and plant growth [53]. UVR8 proteins from green algae to higher plants are functionally conserved and likely to be pivotal in mediating responses to UV-B in numerous species.

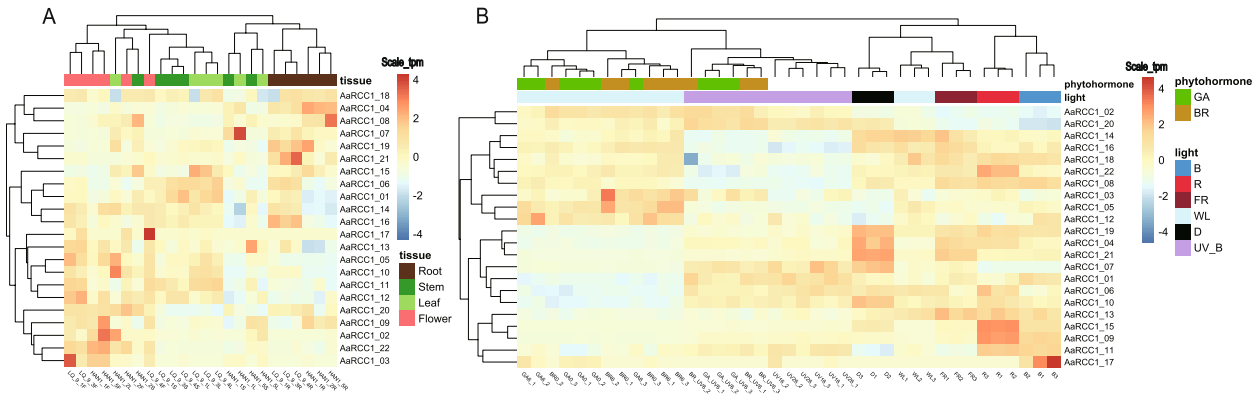

Fig. 3 The expression patterns of *AaRCC1* genes. **A** The expression profile of 22 *AaRCC1* genes in four tissues of two strains. **B** The expression profile of 22 *AaRCC1* genes under different treatments. GA, gibberellin; BR, brassinolide; B, blue light; R, red light; FR, far red light; WL, white light; D, dark; UV_B, UV-B radiation

Strong purifying selection pressure identified among *UVR8* orthologs in different lineages maintains its conserved function. Interestingly, the expression of *AaUVR8* showed a decreasing trend after UV-B treatment, which was different from well-known stress-tolerance genes (NAC, ERF, CBF) [54–56]. It was considered that the UV-B signal was transduced immediately by *UVR8* and relevant genes while repressors (like RUP1 and RUP2) had negative feedback regulation and repressed *UVR8* expression.

Conclusions

In this study, a comprehensive bioinformatic analysis was conducted on the *AaUVR8* and *RCC1* gene family of *A. annua*, which would help explain the role of light signal recognition and transduction in *A. annua.* Besides, the study contributed to screening varieties with high resistance to light stress in molecular-assisted breeding.

Materials and methods

Identification, phylogenetic and conserved domain analysis of the *RCC1* genes in *A. annua*

Four haplotype genomes (LQ-9 haplotype 0, LQ-9 haplotype 1, HAN1 haplotype 0 and HAN1 haplotype 1) and transcriptomes from different tissues of *A. annua* were used in this study, The data were downloaded from Global Pharmacopoeia Genome Database (GPGD, http://www.gpgenome.com/) [50, 57]. The annotated AtUVR8 protein sequences were obtained from the TAIR database (http://www.arabidopsis.org). Protein sequences from *A. annua* genome were searched against the PFAM database (Pfam 32.0) using PfamScan (evalue ≤ 1e-5; http://www.ebi.ac.uk/Tools/pfa/pfamscan). Genes with hits to *RCC1* domain (PF00415) were considered as candidate *RCC1* genes. Finally, the genes were viewed and corrected using the Apollo browser [58] followed the Wang et al. [59] to rule out false-positive results. According to the amino acid similarity (identity ≥ 80%), the allelic (one-to-one) relationship of *RCC1* genes among haplotype genomes was confirmed.

The phylogenetic trees of 22 AaRCC1 proteins in *A. annua* LQ-9 haplotype 0 were constructed using MEGA X [60] with 1000 bootstrap replications and both neighbor-joining and maximum likelihood models. The phylogenetic tree, gene structures and PFAM domains were plotted by ggtree package [61].

Evolutionary analysis of UVR8 in five eudicot species

RCC1 genes of four other species, *A. thaliana*, *C. nankingense*, *H. annuus*, and *V. vinifera* were identified using same method as used for *A. annua*. The *A. annua* RCC1 proteins were searched against candidate RCC1 proteins of other species utilizing BLASTp [62] and hits with identity ≥ 40% and coverage ≥ 60% were kept. The synteny analysis between *A. annua* and other species was performed by the Multiple Collinearity Scan toolkit (MCscan**,** Python version) [63]. The Ks values of ortholog pairs or paralog pairs in five species were calculated using KaKs_Calculator2.0 [64]. Multiple sequence alignment of ortholog proteins was performed using ClustalX method with MEGA X.

Expression analysis based on RNA-Seq data

The raw reads generated by different tissues (http://www.gpgenome.com/species/92) and corresponding transcriptome data with different treatments in A. *annua* (Table S4) deposited in PRJNA435470 (SRP133983) [65] and PRJNA601869 [17] of the NCBI were quality controlled using Skewer [66]. High-quality reads were mapped to the LQ-9 haplotype 0 genome sequences using HISAT2 [67]. The expression level of each gene was calculated with StringTie [68]**.** Differential expression of *RCC1* genes in four tissues (root, stem, leaf, and flower) and different treatments were analyzed with DESeq2 [69]. Hierarchical clustering analysis and expression level of TPM (transcript per million) values was performed using the 'pheatmap' package (https://cran.rproject.org/web/packages/pheatmap/) in R.

RNA extraction and Expression analysis by quantitative PCR

The qPCR samples of *A. annua* were collected in different provinces of China (Table S3), which were identified by Li Xiang and preserved in an accessible herbarium of Artemisinin Research Center, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences. Total RNA was extracted according to the instruction manual of the Plant Total RNA Isolation Kit (Vazyme, Nanjing, China). First-strand cDNA was synthesized with a HiScript III 1st Strand cDNA Synthesis Kit (+gDNA wiper) (Vazyme, Nanjing, China) according to the manufacturer's instructions. *AaActin* was used as a reference. Primers for *AaUVR8* and *AaActin* (Table S2) were designed and synthesized by Sangon Biotech Co., Ltd (Shanghai, China). The qPCR reaction was performed using the Applied Biosystems ABI 7500 PCR System (ABI, United States). The PCR amplification mixture contained 2 μl of cDNA, 10 μl of ChamQ Universal SYBR qPCR Master Mix (Vazyme Biotech Co., Ltd), 0.4 μl of 10 μM forward and reverse primers, and 7.2 μl ddH_2O. The PCR reaction was performed with the initial denaturation step for 30 s at 95 °C; 40 cycles of 10 s at 95 °C and annealing at 60 °C for 30 s. The melting curves (60–95 °C) were used to check the specificity of each qPCR reaction. The standard curves were generated using a twofold dilution gradient of the cDNA. Amplification

efficiencies ($E = 10^{-1/slope}$-1) and correlation coefficient (R^2) values were calculated by standard curves. The relative gene expression was calculated with the $2^{-\Delta\Delta Ct}$ method [70].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-023-09786-4.

Additional file 1: Supplemental Table S1. The alleles of *AaRCC1* in four haplotype genomes. **Supplemental Table S2.** Primers used in qPCR of *AaUVR8* and *AaActin*. **Supplemental Table 3.** qPCR samples of *A. annua*. **Supplemental Table 4.** Different treatments in *A. annua*.

Additional file 2: Fig. S1. The phylogenetic relationships of RCC1 family proteins in 5 species.

Additional file 3: Fig. S2. The duplication debris of *AaRCC1_07* and *AaUVR8* in *A. annua* genome.

Additional file 4: Fig. S3. The relative expression of *AaUVR8* in 18 *A. annua* samples

Additional file 5: Fig. S4. The expression profile of *AtRCC1* genes in UV-B treatment.

Acknowledgements
Not applicable.

Institutional review board statement
Not applicable.

Informed consent statement
Not applicable.

Authors' contributions
Conceptualization, B.L., X.Q. and Z.H.; methodology, J.C., W.W. and B.L.; formal analysis, J.C. and X.D.; investigation, W.W.; resources, W.W., D.Z. and C.D.; data curation, X.D., H.P., C.W., J.Z., J.Z. and P.S.; writing—original draft preparation, J.C.; writing—review and editing, J.C., B.L. and X.Q.; visualization, W.W. and J.C.; supervision, B.L., X.Q. and Z.H.; project admin-istration, B.L., X.Q. and Z.H.; funding acquisition, B.L., Z.H., X.Q. and W.W. All authors have read and agreed to the published version of the manuscript.

Funding
This research was funded by National Natural Science Fundation of China (grant number 82204548 and 81903754), the Fundamental Research Funds for the Central public welfare research institutes (grant number ZZ13-YQ-107), the Young Elite Scientists Sponsorship Program from China Association of Chinese Medicine (grant number CACM-2022-QNRC2-B30) and the Project Quality Standard System Construction for the Whole Industry Chain of Chinese Medicinal Decoction Pieces from Guangdong Provincial Drug Administration of China (grant number 002009/2019KT1261/2020ZDB25).

Availability of data and materials
The identified *RCC1* genes were deposited in the Global Pharmacopoeia Genome Database at http://www.gpgenome.com/species/92. Data supporting the findings of this work are available within the paper and its Supplementary Information files. The datasets generated and analyzed during the study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 27 February 2023 Accepted: 6 November 2023
Published online: 18 November 2023

References

1. Cheong DHJ, Tan DWS, Wong FWS, Tran T. Anti-malarial drug, artemisinin and its derivatives for the treatment of respiratory diseases. Pharmacol Res. 2020;158:104901.
2. Al-Khayri JM, Sudheer WN, Lakshmaiah VV, Mukherjee E, Nizam A, Thiruvengadam M, et al. Biotechnological Approaches for Production of Artemisinin, an Anti-Malarial Drug from *Artemisia annua* L. Molecules 2022;27(9):3040.
3. Tu Y. The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. Nat Med. 2011;17(10):1217–20.
4. Organization WH: World malaria report 2021. In. Edited by WHO; 2021.
5. Bhattarai A, Ali AS, Kachur SP, Martensson A, Abbas AK, Khatib R, et al. Impact of artemisinin-based combination therapy and insecticide-treated nets on malaria burden in Zanzibar. PLoS Med. 2007;4(11):1784–90.
6. Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, Leavell MD, Tai A, Main A, Eng D, et al. High-level semi-synthetic production of the potent antimalarial artemisinin. Nature. 2013;496(7446):528.
7. Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J, et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature. 2006;440(7086):940–3.
8. Peplow M. Synthetic biology's first malaria drug meets market resistance. Nature. 2016;530(7591):389–90.
9. Wang DY, Shi CY, Alamgir K, Kwon S, Pan LL, Zhu YJ, Yang XH. Global assessment of the distribution and conservation status of a key medicinal plant (*Artemisia annua* L.): The roles of climate and anthropogenic activities. Sci Total Environ. 2022;821:14.
10. Li L, Josef BA, Liu B, Zheng SH, Huang LF, Chen SL. Three-Dimensional Evaluation on Ecotypic Diversity of Traditional Chinese Medicine: A Case Study of *Artemisia annua* L. Front Plant Sci. 2017;8:1225.
11. Huang L, Xie C, Duan B, Chen S. Mapping the potential distribution of high artemisinin-yielding *Artemisia annua* L. (Qinghao) in China with a geographic information system. Chin Med. 2010;5:18.
12. Ferreira JFS, Simon JE, Janick J. Developmental studies of *Artemisia annua*: flowering and artemisinin production under greenhouse and field conditions. Planta Medica. 1995;61(2):167–70.
13. Widiyastuti Y, Subositi D. Photoperiod Effect on The Growth and Artemisinin Content of *Artemisia Annua* Grown in Tropical Region. AIP Conference Proceedings. 2019; https://doi.org/10.1063/1.5098432.
14. Björn LO. Ultraviolet-A, B, and C. UV4. Plants Bulletin. 2015; https://doi.org/10.19232/uv4pb.2015.1.12.
15. Frohnmeyer H, Staiger D. Ultraviolet-B radiation-mediated responses in plants. Balancing damage and protection Plant Physiol. 2003;133(4):1420–8.
16. Pandey N, Pandey-Rai S. Short term UV-B radiation-mediated transcriptional responses and altered secondary metabolism of in vitro propagated plantlets of *Artemisia annua* L. Plant Cell Tissue and Organ Culture. 2014;116(3):371–85.
17. Ma TY, Gao H, Zhang D, Shi YH, Zhang TY, Shen XF, et al. Transcriptome analyses revealed the ultraviolet B irradiation and phytohormone gibberellins coordinately promoted the accumulation of artemisinin in *Artemisia annua* L.Chin Med. 2020;15(1):67.
18. Rizzini L, Favory JJ, Cloix C, Faggionato D, O'Hara A, Kaiserli E, et al. Perception of UV-B by the *Arabidopsis* UVR8 Protein. Science. 2011;332(6025):103–6.
19. Jenkins GI. Structure and function of the UV-B photoreceptor UVR8. Curr Opin Struct Biol. 2014;29:52–7.
20. Jenkins GI. The UV-B Photoreceptor UVR8: From Structure to Physiology. Plant Cell. 2014;26(1):21–37.

21. Ulm R, Jenkins GI. Q&A: How do plants sense and respond to UV-B radiation? BMC Biology. 2015;13:45
22. Kliebenstein DJ, Lim JE, Landry LG, Last RL. *Arabidopsis* UVR8 regulates ultraviolet-B signal transduction and tolerance and contains sequence similarity to human Regulator of Chromatin Condensation 1. Plant Physiol. 2002;130(1):234–43.
23. Bischoff FR, Ponstingl H. Catalysis of guanine nucleotide exchange on Ran by the mitotic regulator RCC1. Nature. 1991;354(6348):80–2.
24. Renault L, Nassar N, Vetter I, Becker J, Klebe C, Roth M, Wittinghofer A. The 1.7 Å crystal structure of the regulator of chromosome condensation (RCC1) reveals a seven-bladed propeller. Nature. 1998;392(6671):97–101.
25. Brown BA, Cloix C, Jiang GH, Kaiserli E, Herzyk P, Kliebenstein DJ, et al. A UV-B-specific signaling component orchestrates plant UV protection. Proc Natl Acad Sci USA. 2005;102(50):18225–30.
26. Favory JJ, Stec A, Gruber H, Rizzini L, Oravecz A, Funk M, et al. Interaction of COP1 and UVR8 regulates UV-B-induced photomorphogenesis and stress acclimation in *Arabidopsis*. EMBO J. 2009;28(5):591–601.
27. Kaiserli E, Jenkins GI. UV-B promotes rapid nuclear translocation of the *Arabidopsis* UV-B-specific signaling component UVR8 and activates its function in the nucleus. Plant Cell. 2007;19(8):2662–73.
28. Wu M, Grahn E, Eriksson LA, Strid A. Computational evidence for the role of *Arabidopsis thaliana* UVR8 as UV-B photoreceptor and identification of its chromophore amino acids. J Chem Inf Model. 2011;51(6):1287–95.
29. Takeuchi H. UV Raman Markers for Structural Analysis of Aromatic Side Chains in Proteins. Anal Sci. 2011;27(11):1077–86.
30. Christie JM, Arvai AS, Baxter KJ, Heilmann M, Pratt AJ, O'Hara A, Kelly SM, Hothorn M, Smith BO, Hitomi K, et al. Plant UVR8 photoreceptor senses UV-B by tryptophan-mediated disruption of cross-dimer salt bridges. Science. 2012;335(6075):1492–6.
31. Camacho IS, Theisen A, Johannissen LO, Díaz-Ramos LA, Christie JM, Jenkins GI, et al. Native mass spectrometry reveals the conformational diversity of the UVR8 photoreceptor. Proc Natl Acad Sci USA. 2019;116(4):1116–25.
32. Tossi VE, Regalado JJ, Iannicelli J, Laino LE, Burrieza HP, Escandón AS, et al. Beyond *Arabidopsis*: Differential UV-B Response Mediated by UVR8 in Diverse Species. Front Plant Sci. 2019;10.
33. Li W, Wen J, Song Y, Yuan H, Sun B, Wang R, et al. *SaRCC1*, a Regulator of Chromosome Condensation 1 (RCC1) Family Protein Gene from *Spartina alterniflora*, Negatively Regulates Salinity Stress Tolerance in Transgenic *Arabidopsis*. Int J Mol Sci. 2022;23(15).
34. Liu X, Wu XC, Sun CD, Rong JK. Identification and Expression Profiling of the Regulator of Chromosome Condensation 1 (*RCC1*) Gene Family in *Gossypium Hirsutum* L. under Abiotic Stress and Hormone Treatments. Int J Mol Sci. 2019;20(7):1727.
35. Kuhn K, Carrie C, Giraud E, Wang Y, Meyer EH, Narsai R. des Francs-Small CC, Zhang BT, Murcha MW, Whelan J: The RCC1 family protein RUG3 is required for splicing of nad2 and complex I biogenesis in mitochondria of *Arabidopsis thaliana*.Plant J. 2011;67(6):1067–80.
36. Ji H, Wang Y, Cloix C, Li K, Jenkins GI, Wang S, Shang Z, Shi Y, Yang S, Li X. The *Arabidopsis* RCC1 Family Protein TCF1 Regulates Freezing Tolerance and Cold Acclimation through Modulating Lignin Biosynthesis. PLoS Genet. 2015;11(9):e1005471.
37. Dong ZH, Wang H, Li X, Ji HT. Enhancement of plant cold tolerance by soybean *RCC1* family gene GmTCF1a. BMC Plant Biol. 2021;21(1):369.
38. Furutani M, Hirano Y, Nishimura T, Nakamura M, Taniguchi M, Suzuki K, et al. Polar recruitment of RLD by LAZY1-like protein during gravity signaling in root branch angle control. Nat Commun. 2020;11(1):76.
39. Ji H, Wang S, Cheng C, Li R, Wang Z, Jenkins GI, Kong F, Li X. The RCC1 family protein SAB1 negatively regulates ABI5 through multidimensional mechanisms during postgermination in *Arabidopsis*. New Phytol. 2019;222(2):907–22.
40. Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet. 2002;18(9):486–7.
41. Yang HL, Liu YJ, Wang CL, Zeng QY. Molecular Evolution of Trehalose-6-Phosphate Synthase (TPS) Gene Family in *Populus*, *Arabidopsis* and Rice. PLoS One. 2012;7(8):e42438.
42. Kawai Y, Ono E, Mizutani M. Expansion of specialized metabolism-related superfamily genes via whole genome duplications during angiosperm evolution. Plant Biotechnology. 2014;31(5):579–84.
43. Yao Y, Gu JJ, Luo YJ, Wang YY, Pang YZ, Shen GA, Guo BL. Genome-wide analysis of UGT gene family identified key gene for the biosynthesis of bioactive flavonol glycosides in *Epimedium pubescens* Maxim. Synthetic Syst Biotechnol.. 2022;7(4):1095–107.
44. Jenkins GI. Photomorphogenic responses to ultraviolet-B light. Plant Cell Environ. 2017;40(11):2544–57.
45. Koch MA, Haubold B, Mitchell-Olds T. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). Mol Biol Evol. 2000;17(10):1483–98.
46. Huang CH, Zhang C, Liu M, Hu Y, Gao T, Qi J, Ma H. Multiple Polyploidization Events across Asteraceae with Two Nested Events in the Early History Revealed by Nuclear Phylogenomics. Mol Biol Evol. 2016;33(11):2820–35.
47. Gruber H, Heijde M, Heller W, Albert A, Seidlitz HK, Ulm R. Negative feedback regulation of UV-B-induced photomorphogenesis and stress acclimation in *Arabidopsis*. Proc Natl Acad Sci U S A. 2010;107(46):20132–7.
48. Cen QW, Kang LH, Zhou DN, Zhang X, Tian QX, Zhang XQ, et al. Genome-Wide Identification and Expression Analysis of *RCC1* Gene Family under Abiotic Stresses in Rice (*Oryza sativa* L.). Agronomy-Basel. 2023;13(3):703.
49. An X, Zhao SQ, Luo XH, Chen CL, Liu TT, Li WL, et al. Genome-wide identification and expression analysis of the regulator of chromosome condensation 1 gene family in wheat (*Triticum aestivum* L.). Front Plant Sci. 2023;14:1124905.
50. Liao BS, Shen XF, Xiang L, Guo S, Chen SY, Meng Y, et al. Allele-aware chromosome-level genome assembly of *Artemisia annua* reveals the correlation between ADS expansion and artemisinin yield. Mol Plant. 2022;15(8):1310–28.
51. Gout JF, Hao Y, Johri P, Arnaiz O, Doak TG, Bhullar S, et al. Dynamics of Gene Loss following Ancient Whole-Genome Duplication in the Cryptic *Paramecium* Complex. Mol Biol Evol. 2023;40(5):msad107.
52. Yang Y, Liang T, Zhang LB, Shao K, Gu XX, Shang RX, et al. UVR8 interacts with WRKY36 to regulate HY5 transcription and hypocotyl elongation in *Arabidopsis*. Nat Plants. 2018;4(2):98–107.
53. Liang T, Yang Y, Liu HT. Signal transduction mediated by the plant UV-B photoreceptor UVR8. New Phytol. 2019;221(3):1247–52.
54. Borras D, Barchi L, Schulz K, Moglia A, Acquadro A, Kamranfar I, et al. Transcriptome-Based Identification and Functional Characterization of NAC Transcription Factors Responsive to Drought Stress in *Capsicum annuum* L. Front Genet. 2021;12:743902.
55. Lee CM, Thomashow MF. Photoperiodic regulation of the C-repeat binding factor (CBF) cold acclimation pathway and freezing tolerance in *Arabidopsis thaliana*. Proc Natl Acad Sci U S A. 2012;109(37):15054–9.
56. Shu Y, Liu Y, Zhang J, Song L, Guo C. Genome-Wide Analysis of the AP2/ERF Superfamily Genes and their Responses to Abiotic Stress in *Medicago truncatula*. Front Plant Sci. 2015;6:1247.
57. Liao B, Hu H, Xiao S, Zhou G, Sun W, Chu Y, Meng X, Wei J, Zhang H, Xu J, et al. Global Pharmacopoeia Genome Database is an integrated and mineable genomic database for traditional medicines derived from eight international pharmacopoeias. Science China Life Sciences. 2022;65(4):809–17.
58. Dunn NA, Unni DR, Diesh C, Munoz-Torres M, Harris NL, Yao E, et al. Darling AE. Apollo: democratizing genome annotation. PLoS Comput Biol. 2019;15(2):e1006790.
59. Wang LN, Huang QH, Zhang LL, Wang QF, Liang L, Liao BS. Genome-Wide Characterization and Comparative Analysis of MYB Transcription Factors in Ganoderma Species. G3-Genes Genomes Genetics. 2020;10(8):2653–2660.
60. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X. Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018;35(6):1547–9.
61. Yu G. Using ggtree to Visualize Data on Tree-Like Structures. Curr Protoc Bioinformatics. 2020;69(1):e96.
62. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
63. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. Synteny and collinearity in plant genomes. Science. 2008;320(5875):486–8.
64. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteomics Bioinformatics. 2010;8(1):77–80.
65. Zhang D, Sun W, Shi YH, Wu L, Zhang TY, Xiang L. Red and Blue Light Promote the Accumulation of Artemisinin in *Artemisia annua* L. Molecules. 2018;23(6):1329

66. Jiang HS, Lei R, Ding SW, Zhu SF: Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics. 2014;15.
67. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37(8):907–15.
68. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016;11(9):1650–67.
69. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
70. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25(4):402–8.

Publisher's Note